PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/62, C07K 3/18 .
A61K 37/43, C07K 13/00
C12P 21/02 //, C12N 15/16
A61K 37/24

A1

(11) International Publication Number: WO 92/10576

(43) International Publication Date: 25 June 1992 (25.06.92)

(21) International Application Number: PCT/US91/09152

(22) International Filing Date: 12 December 1991 (12.12.91)

(30) Priority data:
626,727 13 December 1990 (13.12.90) US

(60) Parent Application or Grant
(63) Related by Continuation
US 626,727 (CIP)
Filed on 13 December 1990 (13.12.90)

(71) Applicant *(for all designated States except US)*: THE UPJOHN COMPANY [US/US]; 301 Henrietta Street, Kalamazoo, MI 49001 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)* : KUBIAK, Teresa, M. [PL/US]; 5844 East B Avenue, Richland, MI 49083 (US). SHARMA, Satish, K. [US/US]; 7649 Hampton Oaks, Portage, MI 49002 (US).

(74) Agent: DELUCA, Mark; Corporate Patents & Trademarks, The Upjohn Company, Kalamazoo, MI 49001 (US).

(81) Designated States: AT (European patent), AU, BB, BE (European patent), BF (OAPI patent), BG, BJ (OAPI patent), BR, CA, CF (OAPI patent), CG (OAPI patent), CH (European patent), CI (OAPI patent), CM (OAPI patent), CS, DE (European patent), DK (European patent), ES (European patent), FI, FR (European patent), GA (OAPI patent), GB (European patent), GN (OAPI patent), GR (European patent), HU, IT (European patent), JP, KP, KR, LK, LU (European patent), MC (European patent), MG, ML (OAPI patent), MN, MR (OAPI patent), MW, NL (European patent), NO, PL, RO, SD, SE (European patent), SN (OAPI patent), SU+,TD (OAPI patent), TG (OAPI patent), US.

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: FUSION POLYPEPTIDES

(57) Abstract

A non-naturally-occurring fusion protein comprising an extension peptide portion covalently linked at its C-terminus to the N-terminus of a biologically active portion is disclosed. The extension peptide portion can be removed by DPP IV cleavage. A use of fusion proteins with DPP IV cleavable extension peptide portions in medecinal preparations is disclosed. A method of purifying desired proteins from a mixture containing a fusion protein is disclosed.

+ DESIGNATIONS OF "SU"

Any designation of "SU" has effect in the Russian Federation. It is not yet known whether any such designation has effect in other States of the former Soviet Union.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	ES	Spain	MG	Madagascar
AU	Australia	FI	Finland	ML	Mali
BB	Barbados	FR	France	MN	Mongolia
BE	Belgium	GA	Gabon	MR	Mauritania
BF	Burkina Faso	GB	United Kingdom	MW	Malawi
BG	Bulgaria	GN	Guinea	NL	Netherlands
BJ	Benin	GR	Greece	NO	Norway
BR	Brazil	HU	Hungary	PL	Poland
CA	Canada	IT	Italy	RO	Romania
CF	Central African Republic	JP	Japan	SD	Sudan
CG	Congo	KP	Democratic People's Republic of Korea	SE	Sweden
CH	Switzerland	KR	Republic of Korea	SN	Senegal
CI	Côte d'Ivoire	LI	Liechtenstein	SU+	Soviet Union
CM	Cameroon	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
DE	Germany	MC	Monaco	US	United States of America
DK	Denmark				

FUSION POLYPEPTIDES

FIELD OF THE INVENTION

The present invention relates to non-naturally occurring fusion polypeptides containing N-terminal portions cleavable by dipeptidylpeptidase IV (DPP IV).

BACKGROUND OF THE INVENTION

The techniques of molecular biology, specifically recombinant DNA technology, allow for the production of relatively large quantities of desirable biologically active polypeptides. Furthermore, the genetic information encoding the polypeptides may be modified to produce relatively large quantities of modified polypeptides. Modifications made to the polypeptides are often used to improve their activity or facilitate their production and/or preparation. Accordingly, much effort has been made to determine what modifications are desirable in order to increase, enhance or otherwise alter the biological activity of desired polypeptides. In addition, there is a great deal of work being done to modify desired polypeptides to facilitate their production and purification.

Naturally produced polypeptides are often initially biosynthesized as larger precursors which are then trimmed by a series of proteolytic cleavages to produce the final products. Accordingly, several proteases exist which recognize and cleave specific amino acids and/or amino acid sequences. These proteases participate in a conversion of a precursor protein to the final polypeptide product.

Once such protease is dipeptidylpeptidase IV (DPP IV) (EC 3.4.14.5). DPP IV was first reported in Hopsu-Havu, V.K. and G.G. Glenner, Histo. Chemie 3:197-201 (1966) and has been shown to be present in many mammalian tissues. DPP IV is presently commercially available from Enzyme Systems Products (Dublin, California). DPP IV recognizes specific amino acid sequences on the N-terminus of proteins. Specifically, DPP IV will cleave a dipeptide from the N-terminus when the second amino acid from the N-terminus is proline (Pro), hydroxyproline (Hyp), alanine (Ala), serine (Ser), and threonine (Thr) and any amino acid is at the N-terminus residue position provided if proline or hydroxyproline is not the amino acid residue third from the N-terminus. DPP IV activity is more efficient when proline or alanine is the second amino acid from the N-terminus and is usually most efficient when that position is occupied by proline. The activity of DPP IV in the stepwise cleavage of "PRO" parts of precursors of naturally occurring peptides is widely reported.

Modern technology has made possible the high level production of biologically active proteins. Important polypeptides can be synthesized using peptide synthesizers or in host cells using recombinant DNA technology. Often, biologically active proteins are administered as drugs. Numerous examples exist in which active proteins are used as therapeutics, prophylactics or to enhance or repress traits. Since DPP IV and other proteases degrade proteins, these drugs

are susceptible to degradation. Thus, a problem of using biologically active polypeptides as drugs is that their sustained presence is diminished and they must therefore be administered more frequently.

The rapid developments in recombinant DNA methodology which have allowed the production of polypeptides, proteins, and their analogs in large quantities in a very short period of time have created a need to isolate in highly efficient and predictable manners these proteins from complex mixtures including the total amount of protein produced by the host cells and those in the growth medium. The purification of heterologous polypeptides produced by host cells can be very expensive and can cause denaturation of the protein product itself. An overview of protein purification techniques is provided in the Background Art section of U.S. Patent Number 4,782,137 issued Nov. 1, 1988 to Hopp et al., incorporated herein by reference.

To circumvent the limitations in the art and provide better methods, recombinant DNA technology may be used to provide desired polypeptides in the form of non-naturally occuring proteins which contain a linker peptide that may be used as a ligand or other target for purification means. For example, U.S. patent number 4,782,137 relates to synthesis of a non-naturally occuring peptide containing an antigenic linker peptide. The non-naturally occuring protein can be passed through a column containing immobilized antibodies which bind to the antigenic linker, thus isolating the non-naturally occuring protein. U.S. Patent 4,569,794 relates to a process of purifying non-naturally occuring proteins which contain N-terminal extensions that have an affinity for immobilized metals. The non-naturally occuring proteins bind to immobilized metal ions in a column. One problem with these methods is that the linker peptide is often undesirable and removal of the linker can be difficult.

The present invention relates to non-naturally occurring fusion proteins which comprise a core protein portion and an N-terminal extension which is cleavable by DPP IV. According to the present invention, a non-naturally occuring protein is provided wherein the extension attached to the core protein is not an N-terminal extension that occurs in nature attached to the core peptide; hence, non-naturally occurring fusion protein. The present invention relates to prodrugs which are DPP IV cleavable non-naturally occuring proteins wherein the core protein portion is a biologically active protein. The present invention relates to DPP IV cleavable non-naturally occuring proteins useful in purification methods whereby the N-terminal extension provides a feature or property which facilitates purification of the non-naturally occuring protein.

The present invention provides non-naturally occuring proteins which have N-terminal extensions that are cleavable by DPP IV such that exposure of the non-naturally occuring protein to DPP IV results in conversion of the non-naturally occuring protein to a desirable

protein. When used as a prodrug, the non-naturally occuring protein is processed into a biologically active protein *in vivo* using DPP IV present in the target species. When used in a purification process, non-naturally occuring protein can be purified by using its specifically designed N-terminus as a ligand and then processed with DPP IV to remove the N-terminal extension and liberate a desired protein.

The present invention allows for the production of a desired protein as a non-naturally occuring protein that is later converted to the desired protein when exposed to DPP IV. Prodrugs are converted to drugs over a course of time using the patients' endogenous DPP IV, thereby achieving sustained presence of the active drug in a patient and reducing the frequency of administration. Pure desired proteins can be isolated using the present invention by producing and purifying non-naturally occuring proteins and then processing the non-naturally occuring proteins *in vitro* with DPP IV to produce the desired protein.

SUMMARY OF THE INVENTION

The present invention relates to a non-naturally occurring fusion protein comprising an extension peptide portion covalently linked at its C-terminus to the N-terminus of a core protein portion, said extension peptide portion being of the formula:

$$A\text{-}X\text{-}Y(X'\text{-}Y)_n$$

wherein

A is optional and when present is methionine;

n is 0-20;

X is selected from the group consisting of all naturally occurring amino acid residues;

X' is selected from the group consisting of all naturally occurring amino acid residues except proline and hydroxyproline;

Y is selected from the group consisting of proline, hydroxyproline, alanine, serine and threonine except when n is 0 then Y is selected from the group consisting of alanine, serine and threonine.

The present invention also relates to the use of such non-naturally occuring proteins in medicinal preparations and to a method of purifying desired proteins from a mixture containing such non-naturally occuring proteins and impurities comprising the steps of selectively contacting said non-naturally occuring protein with material which immobilizes said non-naturally occuring protein, removing said impurities, separating said non-naturally occuring proteins from said material, contacting said non-naturally occuring protein with DPP IV, and isolating said desired protein.

INFORMATION DISCLOSURE

U.S. Patent No. 4,569,794 issued February 11, 1986 to Smith et al relates to the process of purifying proteins and compounds useful in such processes. The invention describes

a process of isolating fusion proteins which have biologically active polypeptides at the C-terminal end and an N-terminal extension linker that is a metal ion chelating linker. The fusion peptide has an affinity to immobilized metal ions. Impurities can be removed by passing a mixture containing the fusion protein through a column containing immobilized metal ions. The fusion protein becomes associated with the metal ions and only the impurities are eluted. Upon changing conditions the fusion peptide is liberated from the immobilized metal ions thus resulting in purified fusion protein.

U.S. Patent No. 4,782,137 issued November 1, 1988 to Hopp et al., discloses the synthesis of a fusion protein having a highly antigenic N-terminal portion and a desired polypeptide at the C-terminal portion. According to Hopp et al., the fusion proteins are purified from crude supernatant by passing crude supernatent through a column containing immobilized antibodies which recognize the antigenic portion of the fusion protein. The immobilized antibodies keep the protein in the column while the undesired components of the supernatent are eluted. The column conditions can then be changed to cause the antigen-antibody complex to dissociate. The fusion protein is then eluted and collected.

U.S. Patent No. 4,734,399 issued March 29, 1988, to Felix, et al. relates to growth hormone releasing factor analogs. Several analogs are disclosed which have end terminals of Tyr-Ala and His-Ala. However, these molecules are not fusion proteins but rather core proteins only. The N-terminal dipeptides of Felix, et al. are part of the bGRF analog core molecule.

European patent application Publication Number 0 220 958, published May 6, 1987 relates to selective chemical removal of N-terminal residues. The invention relates to a process and compounds useful in the process to remove N-terminal residues from desired polypeptides. The desired polypeptide exists as a fusion protein having the desired polypeptide link at the N-terminal to a linker having the formula X-Pro. Upon exposure of the fusion protein to specific buffer conditions a diketopiperazine of the X-Pro portion of the fusion protein is formed and cleaved, thereby producing the desired polypeptide from the fusion precursor. The fusion proteins of EPO 220,958 ('958) is not included in the present invention because according to the present invention, when the N-terminal extension is only a dipeptide, i.e., when A is absent, n is zero and X is a naturally-occuring amino acid, Y is either alanine, serine or threonine. Thus, whenever the extension is a dipeptide, it is X-Ala, X-Ser or X-Thr. The '958 application teaches chemical, not enzymatic, cleavage of the dipeptide X-Pro. The dipeptide X-Ala, X-Ser and X-Thr are not susceptible to the type chemical cleavage taught by the '958 application that cuts the X-Pro extension from the core protein.

Australian Patent Application Document No. AU-A-12709/88 discloses fusion proteins which contain affinity peptides useful in immobilized metal affinity chromatography (IMAC).

The affinity peptides disclosed contained at least two neighboring histidine residues. The IMAC purification means disclosed requires a special synthetic chemistry for making nitrilotriacetic acid (NTA) resins.

Tallon, M.A., et al., Biochem. 26:7767-7774 (1987) relate to synthesis of extended analogs of the tridecapeptide α-factor from *Saccharomyces cerevisiae*. The synthesized analogs are extended α-factors, which represent sequences of naturally occurring pro-α-factor coded for in the MFα1 structural gene.

Kriel, G. et al, Eur. J. Biochem. 111:49-58 (1980) describes the stepwise cleavage of the N-terminal portion of melittin precursor (Promelittin) by dipeptidylpeptidase IV. Promelittin is the main constituent of honeybee venom. In the amino acid sequence of the N-terminal portion of the precursor, every second residue is either proline or alanine. When promellitin is exposed to DPP IV isolated from pig kidney, the N-terminal region of the precursor is cleaved in a stepwise fashion producing the mature protein. Promelittin, unlike fusion proteins according to the present invention, is a naturally-occurring protein.

Julius, D., et al, Cell, Vol 32:839-852 (March 1983) relates to the role of membrane bound DPP IV in the processing of yeast α-factor from a larger precursor polypeptide. The yeast α-factor, unlike fusion proteins according to the present invention, is a naturally-occurring protein.

Mollay, C. et al, Eur. J. Biochem. 160:31-35 (1986) describes the isolation of DPP IV from the skin secretion of *Xenopus laevis*. The activity of DPP IV is discussed.

Mentlein, R., FEB, Vol. 234, No. 2, pp. 251-256 (July 1988) reviews proline residues in the maturation and degradation of peptide hormones and neuropeptides. It is reported that in mammals, proline specific proteases such as DPP IV are not involved in the biosynthesis of regulatory peptides but may play an important role in the degradation of them. Thus, it is concluded that while in vertebrates and lower vertebrates precursor proteins rely on DPP IV to convert precursors to mature forms, the processing of regulatory proteins in mammals generally uses DPP IV as a degradation protease.

Frohman, L. A. et al. J. Clin. Invest. 78:906-913 (1986) report that human growth hormone releasing factor (hGRF) and its analogs are rapidly degenerated *in vivo* in humans and *in vitro* by plasma DPP IV.

Frohman, L. A. et al. J. Clin. Invest. 83:1533-1540 (1989) report that human growth hormone releasing factor (hGRF) and its analogs are rapidly degenerated *in vivo* in humans and *in vitro* by plasma DPP IV.

Kubiak, T.M., et al, Drug Metabolism and Disposition, Vol. 17, No. 4, pp. 393-397 (1989) refer to the *in vitro* metabolic degradation of bovine growth hormone releasing factor (bGRF) analogs in bovine and porcine plasma and the correlation with plasma DPP IV activity.

The bGRF analogs tested had an Ala residue at position 2- of the N-terminus. It is reported that the metabolic degradation of bGRF in plasma is due to the presence of DPP IV in the plasma.

Hong, W., et al, Biochemistry, 28:8474-8479 (1989) report the expression of enzymatically active DPP IV in Chinese hamster ovary cells after transfection.

Kreil, G., TIBS 15:23-26 (January 1990) reviews of the stepwise cleavage of dipeptides by DPP's in the conversion of precursors to final products. The precursors, described by Kreil are naturally-occurring proteins. The fusion proteins of the present invention are non-naturally-occurring fusion proteins.

Boman, et al., J. Biol. Chem. 264:5852-5860 (1989) demonstrated that a dipeptidyl peptidase isolated from cecropia pupae (with similar specificity to DPP IV) was able to remove natural N-terminal sequences of Ala-Pro-Glu-Pro from the N-terminal of synthetic copies of the natural precursors of cecropia A and B. The preprocecropin disclosed by Boman is a naturally-occurring protein.

Dalboge, H., et al, Bio/technology, 5:161-164 (February 1987) disclose converting *E. coli* produced precursor of human growth hormone (hGH) to authentic hGH *in vitro*. The N-terminal extension of the precursor is removed by dipeptidypeptidase I.

Dalboge, H., et al, FEBS, Vol. 246 (1,2):89-93 (March 1989) discloses the cloning and expression of IL-1β precursor and its conversion to IL-1β by removal of the precursor's N-terminal extension using dipeptidypeptidase I.

Dalboge, H., et al, FEBS, Vol. 266 (1,2):1-3 (June 1990) refer to in vivo processing of N-terminal methionine in *E. coli*. It is reported that the removal of the N-terminal methionine from extended human growth hormone was dependent upon the amino acid adjacent to the methionine.

Hopp, T.P. et al., Bio/Technol. 6:1204-1210 (October 1988), disclose addition of an eight amino acid peptide to the N-terminus of a desired recombinant lymphokine in order to provide an antigenic N-terminus which can be used in immunoaffinity purification. This publication corresponds to U.S. Patent No. 4,782,137 described above.

Smith, M. C., et al., J. Biol. Chem. Vol. 263, 15:7211-7215 (1988) disclose experimental results supporting the hypothesis that specific metal chelating peptides on the NH_2 terminus of a small peptide can be used to purify that protein using metal ion affinity chromatography. This reference provides specific data regarding one of the examples in the above described U.S. Patent Number 4,569,794. Specifically, the use of the metal chelating peptide His-Trp linked to either luteinizing hormone-releasing hormone or proinsulin allows the chimeric peptide to be purified using IMAC whereas control molecules not containing the His-Trp linker cannot be recovered in the like manner.

Hochuli, E. et al., J. Chromat. 411:177-184 (1987) disclosed a nitrilotriacetic acid absorbent useful for metal chelate affinity chromatography. It is reported that the disclosed absorbent when charged with Ni^{2+} is useful in binding to peptides and proteins containing neighboring histidine residues.

Ljungquist, C. et al., Eur. J. Biochem. 186:563-569 (1989) disclose the use of the metal chelating peptide Ala-His-Gly-His-Arg-Pro in multiplicities of two, four and eight together with a column containing immobilized Zn^{2+} ions. According to Ljungquist use of this metal chelating peptide with zinc columns provides unexpectedly good purification of the fusion proteins.

DETAILED DESCRIPTION OF THE INVENTION

As used herein, the terms "non-naturally occurring fusion protein", "non-naturally occurring fusion polypeptide", "fusion polypeptides" and "fusion proteins" refer interchangeably to proteins and polypeptides which do not normally occur in nature and which comprise a core protein portion and an extension portion.

As used herein "core protein", "core protein portion" and "polypeptide portion" refers to the portion of a fusion polypeptide which is located at the C-terminus end of the molecule and which, absent the extension portion, would be a desired polypeptide and/or a biologically active protein including naturally occuring biologically active proteins and polypeptides and analogs and mutants thereof.

As used herein "N-terminal extension" refers to the first up to about 45 amino acids starting at the N-terminus and which are not part of the core protein.

As used herein "prodrugs" refers to fusion proteins wherein the biologically desired portion is a biologically active protein useful as a drug.

As used herein "biologically active protein" and "biologically active polypeptides" refer to interchangeable proteins and polypeptides which possess biological activity.

As used herein "desired protein" and "desirable protein" refer interchangeable to proteins and polypeptides which are sought in pure form.

As used herein "extension portion" refers to the portion of a fusion protein which is an N-terminal extension and which is not part of the biologically desired portion.

As used herein "DPP IV cleavable N-terminal extension portion" refers to the extension portion of a fusion protein which has an amino acid sequence that can be removed by the stepwise cleavage by DPP IV.

In the Sequence Listing Section, some amino acid residues have been designated Xaa in Seq ID. The following descriptions apply:

In Seq ID No. 3 Xaa^{29} represents C-terminally amidated Argininyl residue.

In Seq ID No. 4 Xaa^{29} represents C-terminally amidated Argininyl residue.

In Seq ID No. 5 Xaa^{29} represents C-terminally amidated Argininyl residue.

In Seq ID No. 14 Xaa^{29} represents C-terminally amidated Argininyl residue.

In Seq ID No. 18 Xaa^{31} represents C-terminally amidated Argininyl residue.

In Seq ID No. 19 Xaa^{33} represents C-terminally amidated Argininyl residue.

In Seq ID No. 20 Xaa^{39} represents C-terminally amidated Argininyl residue.

In Seq ID No. 21 Xaa^{45} represents C-terminally amidated Argininyl residue.

In Seq ID No. 24 Xaa^{27} represents C-terminally amidated Argininyl residue.

In Seq ID No. 25 Xaa^{31} represents C-terminally amidated Argininyl residue.

In Seq ID No. 26 Xaa^{33} represents C-terminally amidated Argininyl residue.

In Seq ID No. 27 Xaa^{35} represents C-terminally amidated Argininyl residue.

In Seq ID No. 28 Xaa^{37} represents C-terminally amidated Argininyl residue.

In Seq ID No. 29 Xaa^{33} represents C-terminally amidated Argininyl residue.

In Seq ID No. 30 Xaa^{35} represents C-terminally amidated Argininyl residue.

In Seq ID No. 31 Xaa^{37} represents C-terminally amidated Argininyl residue.

In Seq ID No. 32 Xaa^{39} represents C-terminally amidated Argininyl residue.

In Seq ID No. 33 Xaa^{45} represents C-terminally amidated Argininyl residue.

In Seq ID No. 34 Xaa^{43} represents C-terminally amidated Argininyl residue.

In Seq ID No. 35 Xaa^{45} represents C-terminally amidated Argininyl residue.

In Seq ID No. 36 Xaa^{31} represents C-terminally amidated Argininyl residue.

In Seq ID No. 37 Xaa^{31} represents C-terminally amidated Argininyl residue.

In Seq ID No. 38 Xaa^{31} represents C-terminally amidated Argininyl residue.

In Seq ID No. 39 Xaa^{31} represents C-terminally amidated Argininyl residue.

In Seq ID No. 40 Xaa^{31} represents C-terminally amidated Argininyl residue.

In Seq ID No. 41 Xaa^{33} represents C-terminally amidated Argininyl residue.

In Seq ID No. 42 Xaa^{31} represents C-terminally amidated Argininyl residue.

In Seq ID No. 43 Xaa^{33} represents C-terminally amidated Argininyl residue.

The present invention relates to improved proteins and polypeptides. According to the present invention, biologically active polypeptides are first produced as fusion proteins which contain the two portions: a first portion which represents the core protein portion; and, a second portion which is an N-terminal extension portion that is covalently linked at its carboxy terminus to the amino terminus of the first portion. The N-terminal extension portion of the fusion polypeptide possesses an amino acid sequence which renders it subject to cleavage by the dipeptidylpeptidase IV (DPP IV).

A fusion protein according to the present invention has the formula:

Extension portion - Core protein portion

wherein "Extension portion" represents a DPP IV cleavable N-terminal extension; " - "

represents a covalent peptide bond; and "core protein portion" represents any desired peptide which is liberated from the Extension portion by DPP IV processing.

The Extension portion of a fusion protein according to the present invention has an amino acid sequence according to the formula:

$$A\text{-}X\text{-}Y(X'\text{-}Y)_n$$

wherein A is optional, and when present is methionine;

n represents the number of sequentially linked X'-Y groups, that number representing from 0 to 20 of such groups, preferably 0 to 10 groups.

X is selected from the group consisting of any naturally occuring amino acid;

Y is selected from the group consisting of proline, alanine, serine, and threonine, except when n = 0, then Y is selected from the group consisting of alanine, serine, and threonine;

X' is selected from the group consisting of any naturally occuring amino acid except proline or hydroxyproline;

According to the formula, when n = 1, there are two Y residues. Further, it is possible to have up to twenty one Y residues and twenty X' residues in a single embodiment. Individual Y residues and X' residues respectively can be any residue of the group from which they are selected. That is, all of the individual Y residues do not have to be the same in a given embodiment. Similarly, in an embodiment with more than one X' residue, each individual X' residue present can be any amino acid residue except proline and hydroxyproline irrespective of what residue any other X' residue may be. Each individual Y and X' residue respectively must conform to the rules for that particular group and all that is necessary is that the various individual residues at the specific positions follow the rules as articulated above.

Fusion proteins in which (A) is present as methionine (Met) represent sequences useful for the production of biologically active proteins by recombinant DNA methods in *E. coli*. The Met sequence present in these precursors usually will be processed by the *E. coli* enzymatic system or some other means which can be performed by a person with ordinary skill in the art. Protein synthesis in *E. coli* is, under normal circumstances, initiated at the translation initiation codon AUG coding for Met. As a consequence, the newly synthesized polypeptides have a methionine residue as their N-terminal amino acid. *E. coli* possesses an enzymatic activity with the capacity to effectively remove N-terminal Met when the Met N-terminal residue is adjacent to an amino acid with a relatively small side chain like Gly, Ala or Ser as well as Pro. Highly specific removal of the N-terminal Met can be accomplished using cyanogen bromide mediated cleavage of Met. However, for that procedure to be successful, the N-terminal Met must be the only Met in the entire protein sequence; otherwise the cleavage will take place after each Met in the sequence. Accordingly, for fusion proteins containing internal Met sequences, the

second amino acid from the N-terminus must be Pro, Gly, Ala or Ser if the Met is to be removed by the *E. coli* enzymatic system.

In addition to fusion polypeptides, the present invention relates to: recombinant DNA molecules which comprise DNA sequences that encode the fusion polypeptides; methods of using the recombinant DNA molecules; methods of using the fusion polypeptides including methods of purifying desired polypeptides and methods of delivering drugs which comprise administering prodrugs that are converted from precursor to biologically active forms by stepwise proteolytic removal of the N-terminal extension in vivo.

Production of fusion polypeptides can be accomplished using standard peptide synthesis or recombinant DNA techniques both well known to those having ordinary skill in the art. Peptide synthesis is the preferred method of making polypeptides which comprise about 50 amino acids or less. For larger molecules, production in host cells using recombinant DNA technology is preferred.

Fusion polypeptides which contain N-terminal portions that are recognized and cleaved by DPP IV are useful and advantageous over unmodified polypeptides comprising only the core protein portion. The present invention describes two areas of particular utility. The first use is to provide fusion polypeptides, termed "prodrugs", which comprise biologically active polypeptides that are useful as drugs covalently linked to DPP IV cleavable N-terminal extensions. These proforms can be converted into biologically active forms upon cleavage by DPP IV in the body of a human or other animal that has been administered the prodrug. Accordingly, the present invention relates to fusion polypeptides useful as prodrugs, use of fusion polypeptides in a medicinal preparation and to a method of delivering biologically active polypeptides to a patient. A second use for fusion polypeptides according to the present invention is in protein purification processes in which the N-terminal extension is the component of the polypeptide which renders it effective in a purification method and which N-terminal extension is then removable by cleavage using DPP IV. Accordingly, the present invention relates to fusion polypeptides useful in purification procedures and to a method of purifying desired polypeptides. These uses serve as examples to illustrate the utility of the present invention and are not meant to limit the invention in any way.

For both uses, the core protein portion of the fusion protein is liberated from the extension portion by DPP IV activity. In the case of fusion proteins used in purification methods, it is undesirable that the core proteins be substrates for DPP IV cleavage. That is, it is preferred that DPP IV not be able to cleave the core protein after the extension portion has been removed. It is often most desirable that when a core protein is a DPP IV substrate, it is delivered as a prodrug. In such cases, the prodrug can result in sustained presence of the core protein since some of the DPP IV found *in vivo* (e.g. in plasma, kidney tissue and liver tissue)

will be used to process N-terminal extensions and, therefore, delay core protein degradation. That is, the extension portion of the fusion protein can act as a substrate for DPP IV and competitive inhibitor, delaying the DPP IV action on the core protein thereby temporarily protecting the core protein.

As used herein, "prodrug" means a fusion protein which contains a DPP IV cleavable N-terminal extension covalently linked to a core protein portion that is a biologically active polypeptide useful as a drug. Prodrugs according to the present invention can be administered as an individual proform or in combination with other compounds. The preferred embodiment is a well defined individual form of a prodrug. In either case, the proforms are processed by naturally occurring DPP IV normally found in the body.

The advantage of administering a prodrug in a medicinal preparation is that it delays activity and/or provides for extended presence of the biologically active protein. Prodrugs can remain active longer than unmodified molecules. Prodrugs can exist in a non-active state until such time elapses that a sufficient portion of the extension portion is degraded and the molecule becomes active. Prodrugs, therefore, can act as a time delayed drug delivery system. Furthermore, different N-terminal extensions are degraded at different rates, depending on their length and the specific residues present in their amino acid sequence. Combinations of different forms of prodrugs having a variety of N-terminal extensions can be provided which can provide a sustained, steady level of active drug in a patient over a course of time. Prodrugs, therefore, can act as a time delayed drug delivery system.

As described above, DPP IV cleaves off a dipeptide from the N-terminus of a polypeptide provided certain residues occupy certain positions. As used herein, "position one" refers to the amino acid residue position at the N-terminus. As used herein, "position two" refers to the amino acid residue position which is immediately adjacent to position one and which is second from the N-terminus. As used herein, "position three" refers to the amino acid residue position which is immediately adjacent to position two and which is third from the N-terminus. The cleavage which will remove the N-terminal dipeptide occurs between position two and position three provided amino acid three is not proline or hydroxyproline and amino acid two is one of five amino acids: proline (Pro), hydroxyproline (Hyp), alanine (Ala), serine (Ser), or threonine (Thr).

DPP IV cleaves the N-terminal residues at a different rate depending upon which of the four amino acid residues is present at position two. In most cases, DPP IV cleaves most efficiently when position two is occupied by Pro and it is next most efficient when position two is occupied by Ala. When position one is occupied by tyrosine, phenylalanine or histidine, DPP IV works at about the same rate when position two is occupied by Pro or Ala. DPP IV is next most efficient when position two is occupied by Ser. It is least efficient when Thr

occupies the second position.

Using this information, a variety of N-terminal extensions can be designed which are processed at different rates. Thus, a medicament can be administered which comprises either a specific prodrug or a combination of prodrug forms. The prodrugs, bearing an assortment of N-terminal extensions, will each be processed at a rate which is dependent upon their amino acid sequences. This combination of prodrug forms can be formulated to comprise a series of prodrugs that are processed into active polypeptides across a spectrum of time.

The length and amino acid sequence residue makeup are controlling factors in the rate of DPP IV cleavage. Extensions containing all or mostly all alternating Y=Pro will be processed the fastest while those containing Y=Thr will be converted the slowest. In addition it is known that dipeptidyl units X-Pro where X is either Glu or Asp are cleaved much slower than their counterparts where X is a neutral or basic amino acid residue. Since extensions can comprise different residues (of the four) at each cleavage position in the extension, an extremely high number of variations and permutations can exist.

Any biologically active polypeptide can be used as a polypeptide drug. PCT patent application number PCT/US90/02923, incorporated herein by reference, PCT patent application number PCT/US91/08248, incorporated herein by reference, and U.S. Patent Application Serial Number 07/368,231, incorporated herein by reference and each disclose bovine growth hormone releasing factor analogs which can be used in a medicinal preparation as a prodrug according to the present invention. Any of the analogs taught in these applications can be used as a core peptide portion of a fusion protein according to the present invention. Fusion proteins comprising such core protein portions linked to extension portions may be produced by those having ordinary skill in the art using well known methods.

Other examples of embodiments of the present invention include hormones, receptors, enzymes, storage proteins and blood proteins. Specific examples include: Vasoactive Intestinal Peptide (VIP); human β-casomorphin; Gastric Inhibitory Peptide (GIP); Gastric Releasing Peptide (GRP); human Peptide HI; human Peptide YY; fragment 7-37 of glucagon-like peptide-1; glucagon-like peptide-2; substance P; Neuropeptide Y; human Pancreatic Polypeptide; insulin-like growth factor-1 (IGF-1); human growth hormone (hGH); bovine growth hormone (bGH); porcine growth hormone (pGH); prolactin (PRL); human, bovine, porcine or ovine growth hormone releasing factor (GRF); interleukin-1β (IL-1β); EGF; IGF-2; glucagon; corticotropin releasing factor (CRF); dynorfin; somatostatin-14; endothelin; transforming growth factor α (TGF-α); transforming growth factor β (TGF-β); interleukin-4; interleukin-6; nerve growth factor (NGF); tumor necrosis factor (TNF); insulin; fibroblast growth factor (FGF); interferon; CD4; and interleukin-2 (IL-2) or their synthetic or biosynthetic analogs. These polypeptides can also be used to form the core protein portion of fusion proteins

according to the present invention. These polypeptides are meant only to serve as examples of embodiments and are not meant to limit the scope of the present invention.

Smaller fusion proteins according to the present invention can be synthesized, for example, by solid-phase methodology utilizing an Applied Biosystems 430A peptide synthesizer (Applied Biosystems, Foster City, California) as described in detail in PCT/US90/02923 and 07/368,231.

For larger molecules, production in host cells using recombinant DNA is preferred. There are several different methods available to one having ordinary skill in the art who wishes to use recombinant DNA technology to produce fusion proteins. Typically, genes encoding desired polypeptides are inserted in expression vectors which are then used to transform or transfect suitable host cells. The inserted gene is then expressed in the host cell and the desired polypeptide is produced. To produce the fusion polypeptides of the present invention in a like manner, an additional DNA sequence is included in the gene insert. Specifically, DNA encoding the N-terminal extension residues is operably linked to the 5′ end of the gene encoding the desired polypeptide. This additional genetic material must be placed downstream from the promoter of the expression vector so that it is under the control of the promoter. Additionally, it must be placed in proper reading frame with the gene so that the expressed protein product includes the N-terminal extension residues covalently linked to the desired polypeptide.

Therefore, to produce fusion proteins according to the present invention using recombinant DNA technology, oligonucleotides must be designed which encode the amino acid sequence of the desired N-terminal extension and these oligonucleotides must be operably inserted upstream of the 5′ end of the gene encoding the core protein portion, generating a chimeric gene. The techniques to make oligonucleotides and the techniques used to producing a chimeric gene are well known to those having ordinary skill in the art.

In addition to the utility of fusion proteins as prodrugs, the present invention relates to the purification and processing of biologically active recombinant polypeptides. The desired biologically active recombinant polypeptides are most preferably produced in a soluble form or secreted from the host. According to the present invention, the extension portion of the fusion protein can be recognized by purification means. The fusion protein is purified from the material present in the secretion media or extraction solution it is contained in and then processed to remove the extension portion from the core protein portion, thus producing purified desired protein. Accordingly, desired proteins most suited for processing as fusion proteins according to the present invention are those biologically active polypeptides which are not themselves substrates for DPP IV cleavage.

In accordance with the present invention, a gene sequence encoding for a desired

protein is isolated, synthesized or otherwise obtained and operably linked to a DNA sequence coding an extension portion. The hybrid gene containing the gene for a desired protein operably linked to a DNA sequence encoding an extension portion is referred to as a chimeric gene.

Methods and materials for preparing chimeric genes and recombinant vectors, transforming or transfecting host cells using the same, replicating the vectors in host cells and expressing biologically active foreign polypeptides and proteins are described in Principles of Gene Manipulation, by Old and Primrose, 2nd edition, 1981 and Sambrook et al., Molecular Cloning, 2nd Edition, Cold Spring Harbor Laboratory Press, NY (1989), both incorporated herein by reference.

The present invention relates to recombinant chimeric genes which encode fusion proteins, expression vectors containing the same, hosts transformed or transfected with these expression vectors, and process for obtaining these genes, expression vectors, and hosts transformed or transfected with said vectors.

The present invention may be used to purify any prokaryotic or eukaroytic protein that can be expressed as the product of recombinant DNA technology in a transformed or transfected host cell. These recombinant protein products include hormones, receptors, enzymes, storage proteins, blood proteins, mutant proteins produced by protein engineering techniques, or synthetic proteins. The desired polypeptides produced may include HIV RNase H, tPA, IL-1, IL-1 receptor, CD4, human nerve growth factor, sCD4-PE40, human respiratory syncytial virus (RSV) FG chimeric glycoprotein (See U.S. Patent Application Serial No. 07/543,780, incorporated herein by reference), EGF, IGF-1, IGF-2, glucagon, corticotropin releasing factor (CRF), dynorfin, endothelin, transforming growth factor α (TGF-α), *Pseudomonus* endotoxin 40 (PE40), transforming growth factor-β (TGF-β), insulin and analogs thereof.

Examples of purification means include IMAC and immunoaffinity. Other purification means which employ the use of extension peptides that can be removed using DPP IV are within the contemplated scope of the present invention.

In one embodiment of the present invention, fusion proteins comprising a biologically active polypeptide portion and an extension portion which is a metal chelating peptide are useful in an immobilized metal affinity chromatography system.

Immobilized Metal Ion Affinity Chromatography (IMAC) for fractionating proteins was first disclosed by Porath, J. et al., Nature 258:598-599 (1975). Porath disclosed derivatizing a resin with iminodiacetic acid (IDA) and chelating metal ions to the IDA-derivatized resin. Porath disclosed proteins could be immobilized in a column which contained immobilized metal ions. It involves attaching a commonly used iminodiacetic acid (IDA) to a matrix followed by

chelating a metal ion to the IDA-containing resin. The proteins bind to the metal ion(s) through functional groups of amino acid residues capable of donating electrons. Potential electron donating amino acid residues are cysteine, histidines, and tryptophan. Proteins interact with metal ions through one or more of these amino acids with electron donating side chains.

Smith et al. discloses in U.S. Patent No. 4,569,794, incorporated by reference herein, that certain amino acids residues are responsible for the binding of the protein to the immobilized metal ions. However, the bound protein can be eluted by lowering the pH or using competitive counter ligands such as imidazole if histidine side chains are involved in the binding. Histidine-containing di- or tripeptides in proteins have been used to show that IMAC is a selective purification technique. Accordingly, Smith et al. discloses using recombinant DNA techniques to produce a fusion protein comprising a metal chelating peptide covalently liked to a desired polypeptide. The metal chelating peptide is an extension portion that is effectively a handle to the desired polypeptide. This handle can be used in protein purification.

Use of IMAC technology with metal chelating peptides having alternating His residues is disclosed in U.S. Patent Application Serial No. 07/506,605, which is incorporated herein by reference. U.S. patent application Serial No. 07/506,605 discloses specific metal chelating peptides which provide unexpectedly superior results in the IMAC purification of a fusion protein when the metal chelating peptide comprises three to six alternating His residues. Following the teachings of U.S. patent application Serial No. 07/506,605 and U.S. Patent No. 4,569,794, it is possible to employ the commonly used IDA resin in IMAC for the purification of fusion proteins having a metal chelating peptide portion with at least three alternating histidine residues which are constituents of DPP IV-recognized sequences. Construction of fusion proteins and their use in an IMAC system is taught by U.S. Patent No. 4,569,794. Construction and use of a metal chelating peptide portion comprising alternating His residues is taught in U.S. Patent Application Serial No. 07/506,605. By providing a fusion protein with DPP IV recognizable residues between alternating His residues the present invention provides a fusion protein which can be purified using IMAC technology and subsequently processed with DPP IV to yield a desired polypeptide.

According to this embodiment of the present invention, the extension portion is a metal chelating peptide which can be represented by the formula:

$$A\text{-}X\text{-}Y(X'\text{-}Y)_n$$

and further, wherein A is optional, and when present is methionine;

n represents the number of sequentially linked X'-Y groups, that number representing from 0 to 20 of such groups, preferably 0 to 10 groups.

X is selected from the group consisting of any naturally occuring amino acid;

Y is selected from the group consisting of proline, alanine, serine, and threonine,

except when n = 0, then Y is selected from the group consisting of alanine, serine, and threonine;

X' is selected from the group consisting of any naturally occuring amino acid except proline or hydroxyproline;

wherein at least two to three residues designated X' and, optionally, X are Histidine (His). Preferably, Y is Pro and n is at least 3. When treated with DPP IV, the N-terminal extension is cleaved in a stepwise fashion, producing the biologically active polypeptide provided the biologically active polypeptide is not itself a DPP IV substrate.

One example of a fusion protein includes an extension portion having the formula

$$\text{His-Y(His-Y)}_n$$

wherein n=3 to 8, and Ys are Pro, Hyp, Ala, Ser or Thr, Pro being the most preferred. Another example of a fusion protein includes an Extension portion having the formula

$$\text{X-Y(His-Y)}_n$$

wherein n is 3 to 8, X is any naturally occurring amino acid; and Ys are Pro, Hyp, Ala, Ser or Thr, Pro being the most preferred.

Since N-terminal Met is a consequence of protein synthesis in *E.coli* and it is known to be processed by the *E. coli* enzymatic system when the adjacent amino acid is Pro, Gly, Ala or Ser, the following extensions represent sequences useful for the IMAC purification and cleavage of biologically active peptides or proteins expressed intracellularly in *E. coli* by recombinant DNA techniques.

$$\text{Met-X-Y(His-Y)}_n$$

wherein n = 3 to 8, X is Pro, Gly, Ser, or Ala; and Ys are Pro, Hyp, Ala, Ser, or Thr, Pro being the most preferred.

In another example, if the peptide or protein desired is to be secreted from a given host after transformation or transfection then the vectors could be designed so as to secrete the protein or polypeptide using an extension portion which facilitate transport, such as:

$$\text{X-Y-(His-Y)}_n$$

wherein n = 3 to 8, X could be any naturally occurring amino acid compatible with the secretion system from a given host and Ys are Pro, Hyp, Ala, Ser, or Thr.

Another protein purification system which uses fusion proteins and which is well suited for DPP IV processing technology is immunoaffinity purification. U.S. Patent No. 4,782,137 issued November 1, 1988 to Hopp et al., incorporated herein by reference, discloses the synthesis of a fusion protein having a highly antigenic N-terminal portion and a desired polypeptide at the C-terminal portion. According to Hopp et al., the fusion proteins are purified from crude supernatent by passing crude supernatent through a column containing

immobilized antibodies which recognize the antigenic portion of the fusion protein. The immobilized antibodies keep the protein in the column while the undesired components of the supernatant are eluted. The column conditions can then be changed to cause the antigen-antibody complex to dissociate.

According to the present invention, the highly antigenic N-terminal portion of the fusion protein is an extension portion which contains DPP IV recognizable residues. After collection as described in the Hopp patent, the fusion protein according to the present invention can be exposed to DPP IV, thereby removing the extension portion. One of ordinary skill in the art could practice the immunoaffinity purification system of Hopp with N-terminal extensions according to the present invention.

The embodiments and examples described herein serve to illustrate the nature of the present invention and are not meant to limit the scope of the invention. Contemplated equivalents include fusion proteins which have N-terminal extensions which can be processed by at least one other means such that removal of the extension is due to a combination of means. Contemplated equivalents also include fusion polypeptides comprising chemically modified amino acid residues.

EXAMPLES

Example 1 Synthetic Prodrugs which are Fusion Prodrugs Having Core Proteins that are DPP IV Substrates

Fusion polypeptides that can be synthesized and administered as prodrugs have a DPP IV degradable N-terminal extension covalently linked to the N-terminal of the biologically active polypeptide. The formula for these prodrugs can be expressed as the formula:

extension portion - core protein drug portion

wherein "extension portion" represents a DPP IV cleavable N-terminal extension; " - " represents a covalent peptide bond; and, "core protein portion" represents any desired peptid which is liberated from the extension portion by DPP IV processing. In this example the core protein of the fusion protein is a potential substrate for DPP IV following removal of the extension portion.

Synthetic prodrugs can be produced using peptide synthesis techniques well known in the art.

In one embodiment, the core protein portion is epidermal growth factor (EGF) and the extension portion is Gly-Pro-Phe-Ala:

$$\text{Gly}^{-4}\text{-Pro}^{-3}\text{-Phe}^{-2}\text{-Ala}^{-1}\text{-[EGF]}.$$

In another embodiment, the core protein portion is glucagon and the extension portion is Ala-Pro-Phe-Ala:

$$\text{Ala}^{-4}\text{-Pro}^{-3}\text{-Phe}^{-2}\text{-Ala}^{-1}\text{-[GLUCAGON]}.$$

In another embodiment, the core protein portion is [Ala^{15} Leu^{27}]-bGRF (1-29)NH_2 (Seq ID 3) and the extension portion is Tyr-Ala:

Tyr^{-2}-Ala^{-1}-{[Ala^{15} Leu^{27}]-bGRF (1-29)NH_2}.

Example 2 Synthetic Prodrugs Which are Fusion Proteins Having Core Proteins that are not DPP IV Substrates

Fusion polypeptides that can be synthesized and administered as prodrugs have a DPP IV degradable N-terminal extension covalently linked to the N-terminal of the biologically active polypeptide. The formula for these prodrugs can be expressed as the formula:

extension portion - core protein drug portion

wherein "extension portion" represents a DPP IV cleavable N-terminal extension; " - " represents a covalent peptide bond; and, "core protein portion" represents any desired peptide which is liberated from the extension portion by DPP IV processing.

Synthetic prodrugs can be produced using peptide synthesis techniques well known in the art.

In one embodiment, the core protein portion is a bGRF analog, [Val^2,$Ser^{8,28}$,Leu^{27}]-bGRF (1-33)OH (Seq ID 1), and the extension portion is Gly-Pro-Tyr-Ala:

Gly^{-4}-Pro^{-3}-Tyr^{-2}-Ala^{-1}-{[$Val^2Ser^{8,28}Ala^{15}Leu^{27}$]-bGRF (1-33)OH}.

In another embodiment, the core protein portion is corticotropin releasing factor (CRF) and the extension portion is Gly-Pro-Phe-Ala:

Gly^{-4}-Pro^{-3}-Phe^{-2}-Ala^{-1}-[CRF].

In another embodiment, the core protein portion is dynorfin and the extension portion is Phe-Pro-Phe-Ala:

Phe^{-4}-Pro^{-3}-Phe^{-2}-Ala^{-1}-[DYNORFIN].

In another embodiment, the core protein portion is somatostatin-28 and the extension portion is Gly-Pro-Phe-Pro:

Gly^{-4}-Pro^{-3}-Phe^{-2}-Pro^{-1}-[SOMATOSTATIN-28].

In another embodiment, the core protein portion is endothelin and the extension portion is Ala-Pro-Phe-Ala:

Ala^{-4}-Pro^{-3}-Phe^{-2}-Ala^{-1}-[ENDOTHELIN].

In another embodiment, the core protein portion is a bGRF analog [$Ile^2Ser^{8,28}Ala^{15}$ Leu^{27}]-bGRF (1-40)OH (Seq ID 2) and the extension portion is Phe-Ala:

Phe^{-2}-Ala^{-1}-{[Ile^2 $Ser^{8,28}$ Ala^{15} Leu^{27}]-bGRF (1-40)OH}.

In another embodiment, the core protein portion is [$Ile^2Ala^{15}Leu^{27}$]-bGRF (1-29) NH_2 (Seq ID 4) and the extension portion is Tyr-Ser:

Tyr^{-2}-Ser^{-1}-{[Ile^2 Ala^{15} Leu^{27}]-bGRF (1-29)NH_2}.

Example 3 Sustained Presence of bGRF Analog Leu^{27}-bGRF (1-29)NH_2

A bGRF analog, Leu^{27}-bGRF (1-29)NH_2, its sequence shown as Seq ID 5, can be administered as a medicament comprising the core protein shown in Seq ID 5 and a variety of N-terminally extended prodrugs.

Several versions of prodrugs can be made by well known methods using Seq ID 5 as the core protein portion. Extension portions for these Seq ID 5-based prodrugs are Ile-Ala, Gly-Pro-Ile-Pro, Seq ID 6, Seq ID 7, Tyr-Ala, Gly-Pro-Tyr-Ala, Seq ID 8, Seq ID 9, Seq ID 10, Seq ID 11, Seq ID 12, Seq ID 13, Tyr-Ala-Tyr-Ala and Val-Ala.

Example 4 Sustained Presence of bGRF Analog [$Thr^2Ala^{15}Leu^{27}$]-bGRF (1-29)NH_2

A bGRF analog, [$Thr^2Ala^{15}Leu^{27}$]-bGRF (1-29)NH_2, its sequences shown as Seq ID 14, can be administered as a medicament comprising the core protein portion shown in Seq ID 14 and a variety of N-terminally extended prodrugs. Three versions of the prodrug were made having extension portions of Tyr-Thr, Tyr-Ser, and Tyr-Thr-Tyr-Thr, respectively.

Example 5 Fusion Proteins which Contain HIV RNase H and N-Terminal Extensions

A strategy to purify chimeric proteins from recombinant E. coli is described based on metal chelating peptide domains containing alternate histidines, with affinity for an immobilized metal ion. Vectors are constructed to direct the synthesis of fusion proteins using HIV RNase H as the core protein. As shown below, these fusion proteins are designed to possess alternating histidines for purification by immobilized metal ion affinity chromatography (IMAC) and alternating prolines or alternating alanines for DPP IV cleavage to remove the metal chelating peptide (mcp).

The preferred DPP IV cleavable N-terminal extensions according to the present invention are outlined as follows:

Fusion protein HIVRH/mcp #1 comprises an N-terminal extension of Seq ID 15 linked to HIV RNase H:

Met^{-11}-Pro^{-10}-Ala^{-9}-His^{-8}-Pro^{-7}-His^{-6}-Pro^{-5}-His^{-4}-Pro^{-3}-His^{-2}-Ala^{-1}-[HIV RNase H]

Fusion protein HIVRH/mcp #2 comprises an N-terminal extension of Seq ID 16 linked to HIV RNase H:

Met^{-11}-Ala^{-10}-Pro^{-9}-His^{-8}-Ala^{-7}-His^{-6}-Ala^{-5}-His^{-4}-Ala^{-3}-His^{-2}-Ala^{-1}-[HIV RNase H]

Fusion protein HIVRH/mcp #3 comprises an N-terminal extension of Seq ID 17 linked to HIV RNase H:

Met^{-11}-Gly^{-10}-Pro^{-9}-His^{-8}-Pro^{-7}-His^{-6}-Pro^{-5}-His^{-4}-Pro^{-3}-His^{-2}-Ala^{-1}-[HIV RNase H]

These fusion proteins are cloned and expressed in *E. coli*, and are purified using DEAE chromatography and RP-HPLC. N-terminal sequencing is used to characterize the fusion

proteins. Application of the alternating histidine-containing fusion proteins to the purification of recombinant proteins by IMAC and subsequent removal of the N-terminal extension by DPP IV confirm the utility of the present invention.

Construction of Chimeric Genes Containing HIV RNase H Gene

All recombinant DNAs are prepared by standard techniques. Oligonucleotides corresponding to the metal chelating peptide/cleavage sequence are constructed, purified, annealed and ligated to a gene encoding HIV RNase H to form a chimeric gene.

To prepare expression vectors encoding alternate histidines/DPP IV recognized cleavage residues/HIV-RNaseH, a chimeric gene is inserted into the final expression vector. Expression vectors containing the chimeric gene constructs are used to transform *E. coli* by standard techniques. Expression of the genes in *E. coli* results in the production of the fusion proteins encoded by the chimeric genes. These fusion proteins contain HIV RNase H amino acids and an N-terminal extension which contains alternate histidines (metal chelating peptide) and alternate prolines or alanines.

Preparation of Crude E. coli Extracts and Isolation of Fusion Proteins for Sequencing

Approximately 3 g of *E. coli* cell paste is suspended in 30 ml of 0.25 M potassium phosphate, pH 7.2 containing 1 mM dithiothreitol (DTT), EDTA, phenylmethylsulfonyl fluoride (PMSF), and benzamidine HCL, 10 mg/liter aprotinin, leupeptin, and bestatin. This suspension is passed through a French Press three times to break the cells. Cell lysates are centrifuged at 12,000 rpm for 1 hr. The supernatant is removed and solid ammonium sulfate added to 70% saturation. After stirring for 1 hr, the suspension is centrifuged at 12,000 rpm for 1 hr. The supernatant is discarded and the pellet is redissolved in 2.25 mls of 50 mM Tris pH 7.5 containing 1 mM DTT, PMSF, and benzamidine. The solution is then dialyzed overnight in 20 mM Tris, 50 mM NaCl, 1 mM DTT, 10% glycerol, and 0.1 mM EDTA pH 7.5 (Buffer A) at 4°C. The dialysate is collected, diluted with one volume of Buffer A, and applied to a 10 ml column of washed DEAE cellulose equilibrated in Buffer A. The run through is collected batchwise and the column further washed with 50 mls of Buffer A. These solutions are collected, pooled, and concentrated by 70% ammonium sulfate precipitation and resuspended in 2 mls of Buffer A and dialyzed as described above. Concentrated RH is used for characterization by N-terminal sequence analysis.

Purification of Fusion Proteins Using IMAC

The feasibility of using a metal chelating peptide for the purification of recombinant proteins from crude extracts can be demonstrated by using the following chimerics expressed in recombinant *E. coli* with HIV RNase H as the model protein. Fusion proteins HIV RNase H/mcp #1, HIV RNase H/mcp #2 and HIV RNase H/mcp #3 are each purified.

IMAC columns are prepared as follows. Chelating Sepharose Fast Flow from

Pharmacia is washed thoroughly with Milli-Q water on a scintered glass filter. The gel is then resuspended in water to form a slurry. The slurry is poured carefully into a glass column (Pharmacia) to a volume of 6 mls (1 x 7 cm). After the gel has settled, the column is washed with 5 volumes of 50 mM EDTA (ethylenediaminetetraacetic acid) pH 8.0. Following this, the column is washed with 5 volumes of 0.2 N NaOH and 5 volumes of Milli-Q water. The column then is charged with 5 volumes of 50 mM $NiSo_4$ (or $ZnCl_2$ or $CuSO_4$). Finally, the column is washed with 5 bed volumes of equilibration buffer. The equilibration buffer is made up of 20 mM Tris pH 8.0, containing 500 mM NaCl, 1 mM PMSF, 1 mM benzamidine, 10 mg/L leupeptin, and 10 mg/L aprotinin.

The column has been equilibrated with at least 5 volumes of equilibration buffer. 5-10 mls of crude recombinant *E. coli* extract are applied to the column by gravity. After all the crude material has entered the column, the column is washed with 10 column volumes of equilibration buffer containing 1.0 M NaCl, instead of 500 mM NaCl, pH 8.0.

The column is then eluted with increasing concentrations of imidazole in the equilibration buffer at pH 8.0. For the earlier experiments, a large number of elutions are performed for each experiment to determine the concentration at which the chimeric eluted. Later this elution is simplified and usually just three imidazole concentrations are used: 35 mM, 100 mM, and 300 mM imidazole in the equilibration buffer, pH 8.0. Ten bed volumes of each imidazole buffer are used. Between elutions, the column is washed with 10 volumes of equilibration buffer. Finally, the column is stripped with 5 bed volumes of 50 mM EDTA pH 8.0 to determine if any protein is still bound to the column. The flow rates for the columns are 1.0 ml/min. 5 ml fractions are collected. The columns are run at room temperature.

Commercially available Pierce protein assay kits are used to determine the protein content of the samples.

HIV RNase H activity is determined by the method described in Becerra, S. P. et al, FEBS 270(1,2):76-80 (September 1990), incorporated herein by reference.

Conversion of the N-terminal extended fusion proteins to mature proteins

Commercially available DPP IV purified from human placenta (Enzyme Systems Products, Dublin, Ca.) with a specific activity of 5200 mU per mg protein is used. One U is equivalent to hydrolysis of 1 umole of a synthetic substrate, Ala-Pro-7-amino-4-2 trifluoromethyl coumarin in one minute at pH 7.8. Enzymatic conversion is carried out by incubation of the fusion protein (about 1-100 mg) at a concentration of 1-10 mg/ml with DPP IV at 25 degrees C for 30 minutes at an enzyme to substrate ratio of 1:100 (w/w). The desired polypeptide is recovered from the uncleaved fusion protein by IMAC. The authenticity is confirmed by N-terminal sequence analysis.

Example 6. Processing of bGRF Analog prodrugs in bovine plasma *in vitro*.

Table 1 summarizes representative experiments to demonstrate generation of the core peptide, [Leu^{27}]-bGRF(1-29)NH_2 (Seq ID 5) from its three N-terminally extended analogs: Tyr^{-4}-Ala^{-3}-Tyr^{-2}-Ala^{-1}-{[Leu^{27}]-bGRF(1-29)NH_2} (Seq ID 19), Ile^{-2}-Ala^{-1}-{[Leu^{27}]-bGRF(1-29)NH_2} (Seq ID 18) and Tyr^{-2}-Ala^{-1}-{[Leu^{27}]-bGRF(1-29)NH_2} (Seq ID 25) upon incubation with bovine plasma *in vitro*. The only metabolites detected in the incubation mixtures were those which were products of DPP-IV-related cleavages.

Tyr^{-4}-Ala^{-3}-Tyr^{-2}-Ala^{-1}-{[Leu^{27}]-bGRF(1-29)NH_2} (Seq ID 19) was sequentially converted over time to Tyr^{-2}-Ala^{-1}-{[Leu^{27}]-bGRF(1-29)NH_2} (Seq ID 25), [Leu^{27}]-bGRF(1-29)NH_2 (core peptide, Seq ID 5) and finally to [Leu^{27}]-bGRF(3-29)NH_2 (Seq. ID 24).

Tyr^{-2}-Ala^{-1}-{[Leu^{27}]-bGRF(1-29)NH_2} (Seq ID 25) was converted to [Leu^{27}]-bGRF(1-29)NH_2 (core peptide, Seq ID 5) and then to [Leu^{27}]-bGRF(3-29)NH_2 (Seq ID 24). The core peptide itself, [Leu^{27}]-bGRF(1-29)NH_2 (Seq ID 5) was converted to [Leu^{27}]-bGRF(3-29)NH_2 (Seq ID 24) by plasma DPP-IV.

Even though no other metabolites were observed under the HPLC conditions used in the experiments, the peptide [Leu^{27}]-bGRF(3-29)NH_2 (Seq ID 24) disappeared over time which indicates that other, non-DPP-IV related, proteolyses must have also been taking place but at significantly slower rates. Not only was the core bGRF peptide generated from the DPP-IV-cleavable bGRF prodrugs shown here but also the half-life of the core protein generated from the fusion protein was significantly prolonged *in vitro* as compared with the half-life of the core peptide (Seq ID 5) incubated directly with bovine plasma (Table 1). Moreover, the time at which the core peptide released from the prodrug is present seems to correlate well with the prodrug extension length: Half-life of Seq ID 5 generated from the prodrug with four amino acid residues in the extension part (Seq ID 19) was longer than that one derived from proGRFs having only two amino acids in the extension like in Seq ID 18 or 25.

Example 7 *In vivo* and *in vitro* bioactivity of bGRF Analog prodrugs.

As shown in Table 2, plasma growth hormone (GH) levels were elevated when Holstein steers were injected iv with analog Seq ID 18. At the dose of 0.2 nmol/kg body weight, the induced growth homone levels in plasma were comparable to those generated upon iv injection with the same dose of the core peptide (Seq ID 5). It is important to stress that the extended peptide Seq ID 18 had only ca 5% inherent potency of the core peptide Seq ID 5 when both were tested in the in vitro pituitary cell assay for GH release. Therefore, the comparable *in vivo* activity of these two peptides seems to indicate that the core peptide could have been released from the extended peptide *in vivo*.

The same pattern of GH release *in vivo* was also observed for the treatment with bGRF prodrugs having four amino acids residues in the extension (Seq ID 19) as shown in Table 3. There was no significant difference in the *in vivo* induced GH release upon treatment with

prodrugs having either 4 or 2 amino acids in the extension (peptides Seq ID 19 and Seq ID 18, respectively) despite the significant difference in the *in vitro* half-life of the core peptide generated from these two bGRF prodrugs *in vitro* as indicated in Table 1. The *in vivo* growth hormone release was rapid and the same for the core peptide Seq ID 5 as well as for Seq ID 19 and 18, with no difference in the time of the GH peak following the challenge with bGRF analogs.

Our interpretation of these results is that most likely the rapid GH release from prodrugs *in vivo* is due to the overall high tissue and organ DPP-IV levels which are ca. 100 fold higher than the plasma DPP-IV concentration. This will explain the difference in the rate of prodrugs processing *in vivo* as compared with the cleavages observed in the in vitro experiments summarized in Table 1. It is also feasible that the half-life of the core peptide generated from its prodrug precursors was extended in vivo but not sufficiently to show an altered (extended) growth hormone release. It is known that GH release *in vivo* is modulated by the stimulatory effect of bGRF and inhibitory action of somatostatin (somatotropin release inhibitory factor, SRIF). In the meal-fed steer model of Moseley, et al. (J. Endocr. 17, 253-259, 1988), animals are injected iv with GRF two hours prior to feeding for the reason that the pituitary is more responsive to a GRF challenge before feeding versus following feeding. Factors associated with feeding such as release of gut/pancreatic SRIF may interfere with the ability of the pituitary to release GH. In other words, no GH will be released from the pituitary even in the presence of GRF during the SRIF overtone. Normally, in the unchallenged meal-fed steer model, serum GH concentration declines to basal levels for 3 to 6 hrs after feeding (so called trough period) with another exogenous episodic GH pulse at 5 to 8 hrs following feeding. In response to GRF injection 2 hrs before feeding, the GH response is rapidly occurring within 5-20 min. and GH level remains elevated for 120 to 240 min. before returning to baseline. In the case of GRF prodrugs tested so far, only the first exogenous GH peak was elevated, the second one did not show any treatment effect. It is possible that the half-life of the core GRF generated from prodrugs in our experiments was not extended long enough to allow for the core peptide to be present in the circulation in sufficient concentrations to affect the second exogenous burst of exogenous GH, usually 4-6 hours after the first one.

Taken together, our results support the general prodrug concept disclosed here because: (i) bGRF prodrugs having DPP-IV-cleavable N-terminal extensions were processed successfully to produce the core peptide(s) in bovine plasma in vitro via DPP-IV mediated cleavages; (ii) the in vitro half-life of the core peptide generated from the prodrugs was significantly longer and was a function of the N-terminal extension length in the prodrug; (iii) the fact that the bGRF prodrugs with very low inherent potency were as effective as the core peptide in the release of GH in vivo indicates that most likely the core peptide was generated in vivo as anticipated.

Example 8 Preparation of Gly^{-4}-Pro^{-3}-Ile^{-2}-Pro^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 26, having the formula: Gly-Pro-Ile-Pro-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.16 (4); Thr 1.07 (1); Ser 1.81 (2); Glu 2.07 (2), Pro 1.98 (2); Gly 1.99 (2); Ala 2.99 (3); Val 1.13 (1), Ile 2.84 (3), Leu 5.08 (5); Tyr 1.93 (2); Phe 0.96 (1); Lys 2.04 (2); Arg 2.97 (3).

Example 9 Preparation of Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Ile^{-2}-Pro^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 27 which comprises Seq ID 6 as the extension portion and which has the formula: #3H-Tyr-Ala-Gly-Pro-Ile-Pro-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.04 (4); Thr 1.03 (1); Ser 1.74 (2); Glu 2.05 (2), Pro 1.99 (2); Gly 2.00 (2); Ala 4.01 (4); Val 1.28 (1), Ile 2.84 (3), Leu 5.09 (5); Tyr 2.94 (3); Phe 0.97 (1); Lys 2.07 (2); Arg 3.00 (3).

Example 10 Preparation of Lys^{-8}-Pro^{-7}-Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Ile^{-2}-Pro^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 28 which comprises Seq ID 7 as the extension portion and which has the formula: Lys-Pro-Tyr-Ala-Gly-Pro-Ile-Pro-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.04 (4); Thr 0.95 (1); Ser 1.78 (2); Glu 2.04 (2), Pro 2.91 (3); Gly 1.98 (2); Ala 3.91 (4); Val 0.96 (1), Ile 2.86 (3), Leu 5.08 (5); Tyr 3.06 (3); Phe 0.97 (1); Lys 3.06 (3); Arg 3.08 (3).

Example 11 Preparation of Gly^{-4}-Pro^{-3}-Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 29 having the formula: #6 Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-

Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.01 (4); Thr 0.96 (1); Ser 1.80 (2); Glu 2.02 (2); Pro 0.97 (1); Gly 1.98 (2); Ala 3.91 (4); Val 0.99 (1), Ile 1.89 (2), Leu 5.08 (5); Tyr 3.05 (3); Phe 0.98 (1); Lys 2.03 (2); Arg 3.06 (3).

Example 12 Preparation of Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Tyr^{2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 30 which comprises Seq ID 8 as the extension portion and which has the formula:

#7 Tyr-Ala-Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.07 (4); Thr 0.96 (1); Ser 1.79 (2); Glu 2.02 (2); Pro 0.99 (1); Gly 1.95 (2); Ala 4.80 (5); Val 0.96 (1), Ile 1.87 (2), Leu 5.09 (5); Tyr 4.11 (4); Phe 0.97 (1); Lys 2.06 (2); Arg 3.08 (3).

Example 13 Preparation of Lys^{-8}-Pro^{-7}-Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 31 which comprises Seq ID 9 as the extension portion and which has the formula:

#8 Lys-Pro-Tyr-Ala-Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.06 (4); Thr 0.95 (1); Ser 1.78 (2); Glu 2.01 (2); Pro 1.95 (2); Gly 1.96 (2); Ala 4.81 (5); Val 0.95 (1), Ile 1.87 (2), Leu 5.09 (5); Tyr 4.12 (4); Phe 0.96 (1); Lys 3.08 (3); Arg 3.10 (3).

Example 14 Preparation of Phe^{-10}-Ala^{-9}-Lys^{-8}-Pro^{-7}-Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 32 which comprises Seq ID 10 as the extension portion and which has the formula:

#9 Phe-Ala-Lys-Pro-Tyr-Ala-Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino

acid analysis, theoretical values in parantheses: Asp 4.16 (4); Thr 1.01 (1); Ser 1.89 (2); Glu 2.08 (2); Pro 1.91 (2); Gly 1.93 (2); Ala 5.72 (6); Val 0.97 (1), Ile 1.90 (2), Leu 5.09 (5); Tyr 4.09 (4); Phe 1.99 (2); Lys 3.08 (3); Arg 3.04 (3).

Example 15 Preparation of Gly^{-12}-Pro^{-11}-Phe^{-10}-Ala^{-9}-Lys^{-8}-Pro^{-7}-Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 33 which comprises Seq ID 11 as the extension portion and which has the formula:

#10 Gly-Pro-Phe-Ala-Lys-Pro-Tyr-Ala-Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.08 (4); Thr 0.96 (1); Ser 1.79 (2); Glu 2.07 (2); Pro 2.88 (3); Gly 2.94 (3); Ala 5.74 (6); Val 0.96 (1), Ile 1.88 (2), Leu 5.13 (5); Tyr 4.11 (4); Phe 1.99 (2); Lys 3.10 (3); Arg 3.09 (3).

Example 16 Preparation of Val^{-14}-Pro^{-13}-Gly^{-12}-Pro^{-11}-Phe^{-10}-Ala^{-9}-Lys^{-8}-Pro^{-7}-Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 34 which comprises Seq ID 12 as the extension portion and which has the formula:

#11 Val-Pro-Gly-Pro-Phe-Ala-Lys-Pro-Tyr-Ala-Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.04 (4); Thr 0.96 (1); Ser 1.81 (2); Glu 2.04 (2); Pro 3.80 (4); Gly 2.99 (3); Ala 5.92 (6); Val 1.98 (2), Ile 1.89 (2), Leu 5.10 (5); Tyr 4.09 (4); Phe 1.99 (2); Lys 3.06 (3); Arg 3.08 (3).

Example 17 Preparation of Arg^{-16}-Pro^{-15}-Val^{-14}-Pro^{-13}-Gly^{-12}-Pro^{-11}-Phe^{-10}-Ala^{-9}-Lys^{-8}-Pro^{-7}-Tyr^{-6}-Ala^{-5}-Gly^{-4}-Pro^{-3}-Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 35 which comprises Seq ID 13 as the extension portion and which has the formula:

#12 Arg-Pro-Val-Pro-Gly-Pro-Phe-Ala-Lys-Pro-Tyr-Ala-Gly-Pro-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.10 (4); Thr 0.98 (1); Ser 1.84 (2); Glu 2.03 (2); Pro 4.82 (5); Gly 2.97 (3); Ala 5.91 (6); Val 1.99 (2), Ile 1.88

(2), Leu 5.09 (5); Tyr 4.10 (4); Phe 1.98 (2); Lys 3.03 (3); Arg 4.09 (4).

Example 18 Preparation of Val^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 36 having the formula: #14 Val-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 3.98 (4); Thr 0.89 (1); Ser 1.76 (2); Glu 2.02 (2); Gly 1.05 (1); Ala 3.87 (4); Val 1.85 (2), Ile 1.77 (2), Leu 5.17 (5); Tyr 2.04 (2); Phe 0.97 (1); Lys 2.07 (2); Arg 3.06 (3).

Example 19 Preparation of Tyr^{-2}-Thr^{-1} {[Ala^{15} Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 37 having the formula: #15 Tyr-Thr-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Ala-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.06 (4); Thr 1.86 (2); Ser 1.77 (2); Glu 2.07 (2), Ala 3.98 (4); Val 1.08 (1), Ile 1.89 (2), Leu 5.14 (5); Tyr 2.94 (3); Phe 0.96 (1); Lys 1.99 (2); Arg 3.04 (3).

Example 20 Preparation of Tyr^{-2}-Thr^{-1} {[Ile^2 Ala^{15} Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 38 having the formula: #16 Tyr-Thr-Tyr-Ile-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Ala-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.07 (4); Thr 1.87 (2); Ser 1.75 (2); Glu 2.07 (2), Ala 2.94 (3); Val 1.09 (1), Ile 2.87 (3), Leu 5.12 (5); Tyr 2.92 (3); Phe 0.96 (1); Lys 2.00 (2); Arg 3.05 (3).

Example 21 Preparation of Tyr^{-2}-Thr^{-1} {[Thr^2 Ala^{15} Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 39 having the formula: #17 Tyr-Thr-Tyr-Thr-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Ala-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.05 (4); Thr 2.68 (3); Ser 1.77 (2); Glu 2.07 (2), Ala 2.90 (3); Val 1.08 (1),

Ile 1.89 (2), Leu 5.20 (5); Tyr 2.87 (3); Phe 0.93 (1); Lys 2.01 (2); Arg 3.07 (3).

Example 22 Preparation of Tyr^{-2}-Ser^{-1} {[Thr^2 Ala^{15} Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 40 having the formula: #18 Tyr-Ser-Tyr-Thr-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Ala-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.11 (4); Thr 1.82 (2); Ser 2.64 (3); Glu 2.05 (2), Ala 2.90 (3); Val 1.04 (1), Ile 1.87 (2), Leu 5.16 (5); Tyr 2.92 (3); Phe 0.94 (1); Lys 2.01 (2); Arg 3.04 (3).

Example 23 Preparation of Tyr^{-4}-Thr^{-3}-Tyr^{-2}-Thr^{-1} {[Thr^2 Ala^{15} Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 41 having the formula: #19 Tyr-Thr-Tyr-Thr-Tyr-Thr-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Ala-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.06 (4); Thr 3.66 (4); Ser 1.85 (2); Glu 2.05 (2), Ala 2.93 (3); Val 1.09 (1), Ile 1.91 (2), Leu 5.15 (5); Tyr 3.91 (4); Phe 0.95 (1); Lys 2.00 (2); Arg 3.04 (3).

Example 24 Preparation of Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 42 having the formula: Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in parantheses: Asp 4.01 (4); Thr 0.97 (1); Ser 1.88 (2); Glu 2.00 (2); Gly 1.02 (1); Ala 3.88 (4); Val 0.97 (1), Ile 1.86 (2), Leu 5.03 (5); Tyr 3.03 (3); Phe 0.96 (1); Lys 2.18 (2); Arg 3.03 (3).

Example 25 Preparation of Tyr^{-4}-Ala^{-3}-Tyr^{-2}-Ala^{-1} {[Leu^{27}] bGRF(1-29)NH_2}, trifluoroacetate salt.

The synthesis of the GRF analog peptide Seq ID 43 having the formula: #13 Tyr-Ala-Tyr-Ala-Tyr-Ala-Asp-Ala-Ile-Phe-Thr-Asn-Ser-Tyr-Arg-Lys-Val-Leu-Gly-Gln-Leu-Ser-Ala-Arg-Lys-Leu-Leu-Gln-Asp-Ile-Leu-Asn-Arg-NH_2 (as the CF_3COOH salt) is conducted in a stepwise manner as in procedure A which is described in published PCT patent application PCT/US90/02923 incorporated herein by reference. Amino acid analysis, theoretical values in

parantheses: Asp 3.97 (4); Thr 0.90 (1); Ser 1.74 (2); Glu 1.98 (2); Gly 1.04 (1); Ala 4.85 (5); Val 0.91 (1), Ile 1.77 (2), Leu 5.13 (5); Tyr 4.14 (4); Phe 0.99 (1); Lys 2.07 (2); Arg 3.05 (3).

Table 1. *In vitro* potency and in vitro plasma stability of selected GRF analogs.

Peptide Sequence	Seq. ID No.	*In Vitro* Potency*	*In Vitro* plasma $t_{1/2}$** (min.)
$[Leu^{27}]$-bGRF(1-29)NH_2	5	1.00	24.8 (Exp 1)
Ile^{-2}-Pro^{-1}{$[Leu^{27}]$-bGRF(1-29)NH_2}	18	0.045	43.2 (Exp 1)
Tyr^{-2}-Ala^{-1}{$[Leu^{27}]$-bGRF(1-29)NH_2}	25	0.13	38.5 (Exp 1)
Tyr^{-4}-Ala^{-3}-Tyr^{-2}-Ala^{-1}{$[Leu^{27}]$-bGRF(1-29)NH_2}	19	0.052	297.1# (Exp 2)

* Peptides were tested in an *in vitro* bovine anterior pituitary cell culture as described by Friedman et al. (Int. J. Peptide and Protein Res. 37:14-20 [1991]).

** Peptides were incubated at 30 μM in bovine plasma *in vitro* at 37°C as described in Kubiak et al. (Drug Met. Disp. 17:393-397 [1989]). Values presented here relate to the half-life of $[Leu^{27}]$-bGRF(1-29)NH_2 (Seq. ID 5) incubated directly in plasma or $[Leu^{27}]$-bGRF(1-29)NH_2 (Seq. ID 5) generated from extended peptides Seq. ID No. 18, 25 or 19, respectively. Two experiments with two different plasma pools were run, Exp 1 and 2 as indicated in parentheses.

\# Peptide Seq. ID 19 was tested against $[Leu^{27}]$-bGRF(1-29)NH_2 (Seq. ID 5) using a different bovine plasma pool. The half-life of Seq. ID 5 in this plasma specimen was 50.2 min.

Table 2. Serum GH Response to IV Injections of Various Doses of [Leu^{27}]-bGRF(1-29)NH_2 (Seq ID 5) and Ile^{-2}-Pro^{-1}{[Leu^{27}]-bGRF(1-29)NH_2} (Seq ID 18) in Meal-Fed Holstein Steers.*

Treatment	Dose nmol/kg	Number of Animals Respond-ing	Peak Height (ng/ml)		Time to Peak (min)		Area 0-8 h (Unit)	
			A^{a}	B^{a}	A^{a}	B^{a}	A^{a}	B^{a}
Saline	0	0^{b}	32.4^{b}	32.4^{b}	89^{b}	89^{b}	4.3^{b}	4.3^{b}
Seq ID 18	0.02	$8/10^{c}$	$71.2^{b,c}$	$76.9^{b,c}$	23^{c}	18^{c}	$4.6^{b,c}$	$5.0^{b,c}$
Seq ID 5	0.20	$9/10^{c}$	119.8^{c}	130.9^{c}	23^{c}	14^{c}	6.8^{d}	7.0^{d}
Seq ID 18	0.20	$10/10^{c}$	$101.4^{c,d}$	101.4^{c}	26^{c}	26^{c}	$6.3^{c,d}$	$6.3^{c,d}$
Seq ID 18	20.0	$10/10^{c}$	137.8^{d}	137.8^{c}	23^{c}	23^{c}	10.1^{e}	10.1^{e}
SEM		.04	9.1	9.4	8	8	.3	.3
p Value		.0001	.007	.007	.04	.03	.0001	.0001

* Animals were injected IV with peptides at the doses indicated 2 hrs before feeding and procedures were as described by Moseley et al. J. Endocrinology 117:253-259 (1988).

a Analysis A includes all steers and Analysis B includes only steers responding to GRF injection and control steers.

b,c,d,e Values with different superscripts in a column are significantly different ($P<.05$).

Table 3. Serum GH Response to IV Injections of Various Doses of [Leu^{27}]-bGRF(1-29)NH_2 (Seq ID 5) and Tyr^{-4}-Ala^{-3}-Tyr^{-2}-Ala^{-1}{[Leu^{27}]-bGRF(1-29)NH_2} (Seq ID 19) in Meal-Fed Holstein Steers.*

Treat-ment	Dose nmol/kg	Number of Animals Respond-ing	Peak Height (ng/ml)		Time to Peak (min)		Area 0-10 h (Unit)	
			A[a]	B[a]	A	B	A	B
Saline	0	---	30.9^{b}	30.9^{b}	45^{b}	45^{b}	3.9^{b}	3.9^{b}
Seq ID 5	0.02	$9/12^{c,d}$	91.9^{c}	101.3^{c}	41^{b}	20^{c}	$4.7^{b,c}$	$4.7^{b,c}$
Seq ID 19	0.02	$10/12^{c,d}$	88.5^{c}	92.4^{c}	30^{b}	22^{c}	4.8^{c}	$4.4^{b,c}$
Seq ID 5	0.20	$11/12^{d}$	79.3^{c}	$79.1^{b,c}$	21^{b}	18^{c}	5.4^{c}	5.3^{c}
Seq ID 19	0.20	$7/12^{c}$	97.4^{c}	120.1^{c}	63^{b}	8^{c}	5.4^{c}	$5.4^{c,d}$
Seq ID 19	2.0	$9/12^{c,d}$	74.5^{c}	$70.8^{b,c}$	24^{b}	18^{c}	6.9^{d}	6.8^{d}
SEM		.10	15.9	17.3	14	8	.4	.4
p Value		.0001	.06	.06	.28	.11	.0008	.007
EMS		.1117	3042	3623	2260	763	2281	2483

* Steers were injected IV with peptides at the doses indicated 2 hrs before feeding and procedures were as described by Moseley et al. J. Endocrinology 117:253-259 (1988).

a Analysis A includes all steers and Analysis B includes only steers responding to GRF injection and control steers.

b,c,d Values with different superscripts in a column are significantly different ($P<.05$).

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Kubiak, Teresa M.
Sharma, Satish K.

(ii) TITLE OF INVENTION: Fusion Polypeptides

(iii) NUMBER OF SEQUENCES: 42

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Upjohn Company - Corp. Patents & Trademarks
(B) STREET: 301 Henrietta Street
(C) CITY: Kalamazoo
(D) STATE: Michigan
(E) COUNTRY: USA
(F) ZIP: 49001

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: diskette (3M 3.5, DS double side 1.0 MB)
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: WordPerfect 5.1

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US07/626,727
(B) FILING DATE: 13/12/90

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US07/614,170
(B) FILING DATE: 14/11/90

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: PCT/US90/02923
(B) FILING DATE: 30/05/90

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US07/368,231
(B) FILING DATE: 16/06/89

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US07/506,605
(B) FILING DATE: 09/04/90

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: DeLuca, Mark
(B) REGISTRATION NUMBER: 33229
(C) REFERENCE/DOCKET NUMBER: 4595

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 616 385 5210
(B) TELEFAX: 616 385 6897

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 33

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Tyr Val Asp Ala Ile Phe Thr Ser Ser Tyr Arg Lys Val Leu Ala Gln
1 5 10 15

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Ser Arg Gln Gln Gly
20 25 30

Glu

(3) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 40

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Tyr Ile Asp Ala Ile Phe Thr Ser Ser Tyr Arg Lys Val Leu Ala Gln
1 5 10 15

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Ser Arg Gln Gln Gly
20 25 30

Glu Arg Asn Gln Glu Gln Gly Ala
35 40

(4) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 29

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Ala Gln
1 5 10 15

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25

(5) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 29

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Tyr Ile Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Ala Gln
1 5 10 15

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25

(6) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 29

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln
1 5 10 15

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25

(7) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 6

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Tyr Ala Gly Pro Ile Pro

1 . 5

(8) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 8

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Lys Pro Tyr Ala Gly Pro Ile Pro
1 5

(9) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 6

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Tyr Ala Gly Pro Tyr Ala
1 5

(10) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 8

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Lys Pro Tyr Ala Gly Pro Tyr Ala
1 5

(11) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 10

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
1 5 10

(12) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 12

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```
Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
 1               5                  10
```

(13) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 14

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```
Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
 1               5                  10
```

(14) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 16

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```
Arg Pro Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
 1               5                  10                  15
```

(15) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 29

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa29

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

```
Tyr Thr Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Ala Gln
 1               5                  10                  15

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
```

20 25

(16) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 11

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Met Pro Ala His Pro His Pro His Pro His Ala
1 5 10

(17) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 11

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

Met Ala Pro His Ala His Ala His Ala His Ala
1 5 10

(18) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 11

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Met Gly Pro His Pro His Pro His Pro His Ala
1 5 10

(19) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 31

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

Ile Pro Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu
1 5 10 15

Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25 30

(20) INFORMATION FOR SEQ ID NO:19:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 33
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa33
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

Tyr Ala Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Ser Ser Tyr Arg Lys
1 5 10 15

Val Leu Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Ser
20 25 30

Xaa

(21) INFORMATION FOR SEQ ID NO:20:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 39
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa39
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala Tyr Ala Asp Ala Ile Phe
1 5 10 15

Thr Asn Ser Tyr Arg Lys Val Leu Ala Gln Leu Ser Ala Arg Lys Leu
20 25 30

Leu Gln Asp Ile Leu Asn Xaa
35

(22) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 45

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

```
Arg Pro Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
 1               5                  10                  15

Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln
            20                  25                  30

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
        35                  40                  45
```

(23) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 10

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```
Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
 1               5                  10
```

(24) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 16

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```
Arg Pro Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
 1               5                  10                  15
```

(25) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 27

(B) TYPE: amino acid

(D) TOPOLOGY: lin ar

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa27

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

```
Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln Leu Ser
 1               5                  10                  15

Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
            20                  25
```

(26) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 31

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

```
Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Ser Ser Tyr Arg Lys Val Leu
 1               5                  10                  15

Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
            20                  25                  30
```

(27) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 33

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

```
Gly Pro Ile Pro Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys
 1               5                  10                  15

Val Leu Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn
            20                  25                  30
```

Xaa .

(28) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 35

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Tyr Ala Gly Pro Ile Pro Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr
1 5 10 15

Arg Lys Val Leu Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile
20 25 30

Leu Asn Xaa
35

(29) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 37

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa37

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

Lys Pro Tyr Ala Gly Pro Ile Pro Tyr Ala Asp Ala Ile Phe Thr Asn
1 5 10 15

Ser Tyr Arg Lys Val Leu Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln
20 25 30

Asp Ile Leu Asn Xaa
35

(30) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 33

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa33
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

```
Gly Pro Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys
 1               5                  10                  15

Val Leu Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn
            20                  25                  30

Xaa
```

(31) INFORMATION FOR SEQ ID NO:30:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 35
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa35
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

```
Tyr Ala Gly Pro Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr
 1               5                  10                  15

Arg Lys Val Leu Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile
            20                  25                  30

Leu Asn Xaa
        35
```

(32) INFORMATION FOR SEQ ID NO:31:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 37
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa37
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

```
Lys Pro Tyr Ala Gly Pro Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Asn
```

1 5 10 15

Ser Tyr Arg Lys Val Leu Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln
20 25 30

Asp Ile Leu Asn Xaa
35

(33) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 39

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa39

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala Tyr Ala Asp Ala Ile Phe
1 5 10 15

Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln Leu Ser Ala Arg Lys Leu
20 25 30

Leu Gln Asp Ile Leu Asn Xaa
35

(34) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 41

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa41

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala Tyr Ala Asp Ala
1 5 10 15

Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln Leu Ser Ala Arg
20 25 30

Lys Leu Leu Gln Asp Il Leu Asn Xaa

35 40

(35) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 43

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa43

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

```
Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala Tyr Ala
 1               5                  10                  15

Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln Leu Ser
            20                  25                  30

Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
        35                  40
```

(36) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 45

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

```
Arg Pro Val Pro Gly Pro Phe Ala Lys Pro Tyr Ala Gly Pro Tyr Ala
 1               5                  10                  15

Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu Gly Gln
            20                  25                  30

Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
        35                  40                  45
```

(37) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 31

(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa31
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

Val Ala Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu
1 5 10 15

Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25 30

(38) INFORMATION FOR SEQ ID NO:37:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 31
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa31
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Tyr Thr Tyr Ala Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu
1 5 10 15

Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25 30

(39) INFORMATION FOR SEQ ID NO:38:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 31
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa31
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

Tyr Thr Tyr Ile Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu
1 5 10 15

Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25 30

(40) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 31

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

```
Tyr Thr Tyr Thr Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu
 1               5                  10                  15

Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
            20                  25                  30
```

(41) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 31

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa31

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

```
Tyr Ser Tyr Thr Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys Val Leu
 1               5                  10                  15

Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
            20                  25                  30
```

(42) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTIC:

(A) LENGTH: 33

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: C-terminally amidated Argininyl residue

(B) LOCATION: Xaa33

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

```
Tyr Thr Tyr Thr Tyr Thr Asp Ala Ile Phe Thr Asn Ser Tyr Arg Lys
```

1 5 10 15

Val Leu Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn
20 25 30

Xaa

(43) INFORMATION FOR SEQ ID NO:42:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 31
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa31
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Ser Ser Tyr Arg Lys Val Leu
1 5 10 15

Gly Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Asn Xaa
20 25 30

(44) INFORMATION FOR SEQ ID NO:43:
(i) SEQUENCE CHARACTERISTIC:
(A) LENGTH: 33
(B) TYPE: amino acid
(D) TOPOLOGY: linear
(ix) FEATURE:
(A) NAME/KEY: C-terminally amidated Argininyl residue
(B) LOCATION: Xaa33
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Tyr Ala Tyr Ala Tyr Ala Asp Ala Ile Phe Thr Ser Ser Tyr Arg Lys
1 5 10 15

Val Leu Ala Gln Leu Ser Ala Arg Lys Leu Leu Gln Asp Ile Leu Ser
20 25 30

Xaa

CLAIMS

1. A non-naturally-occurring fusion protein comprising an extension peptide portion covalently linked at its C-terminus to the N-terminus of a core protein portion, said extension peptide portion being of the formula:

$$A\text{-}X\text{-}Y(X'\text{-}Y)_n$$

wherein

A is optional and when present is methionine;

n is 0-20;

X is selected from the group consisting of all naturally occurring amino acid residues;

X' is selected from the group consisting of all naturally occurring amino acid residues except proline and hydroxyproline;

Y is selected from the group consisting of proline, hydoxyproline, alanine, serine and threonine except when n is zero and A is absent then Y is selected from the group consisting of alanine, serine and threonine.

2. A non-naturally-occurring fusion protein according to claim 1 wherein A is present and X is selected from the group consisting of Pro, Gly, Ala and Ser.

3. A non-naturally-occurring fusion protein according to Claim 1 wherein n is 0-10.

4. A non-naturally-occurring fusion protein according to claim 1 wherein said biologically active polypeptide is selected from the group consisting of: bGRF analogs, EGF; IGF-2, glucagon; corticotropin releasing factor; dynorfin, somatostatin-14; endothelin; transforming growth factor α; Vasoactive Intestinal Peptide; human β-casomorphin; Gastric Inhibitory Peptide; Gastric Releasing Peptide; human Peptide HI; human Peptide YY; glucagon-like peptide-1 fragment 7-37; glucagon-like peptide-2; substance P; Neuropeptide Y; human Pancreatic Polypeptide; insulin-like growth factor-1; human growth hormone; bovine growth hormone; porcine growth hormone; prolactin; human growth hormone releasing factor; bovine growth hormone releasing factor; porcine growth hormone releasing factor; ovine growth hormone releasing factor; interleukin -1β; and interleukin-2.

5. A non-naturally-occurring fusion protein according to claim 1 wherein said extension peptide portion is selected from the group consisting of Gly-Pro-Ile-Pro, Seq ID 6, Seq ID 7, Tyr-Ala, Gly-Pro-Tyr-Ala, Seq ID 8, Seq ID 9, Seq ID 10, Seq ID 11, Seq ID 12, Seq ID 13, Tyr-Ala-Tyr-Ala, Val-Ala, Seq ID 15, Seq ID 16, Seq ID 17, Seq ID 22 and Seq ID 23.

6. A non-naturally-occurring fusion protein according to claim 2 wherein n is 3-5.

7. A non-naturally-occurring fusion protein according to claim 6 wherein all X′ residues are histidines.

8. A non-naturally-occurring fusion protein according to claim 7 wherein X is a histidine.

9. A non-naturally-occurring fusion protein according to claim 7 wherein three Y residues are proline.

10. Use of a non-naturally-occurring fusion protein according to claim 1 to prepare a medicament.

11. A use according to claim 10, wherein said medicament comprises additional fusion proteins having identical biologically active portions and different extension portions.

12. A use according to claim 10 wherein said biologically active portion of said non-naturally-occurring fusion protein is a bGRF analog.

13. A method of purifying desired proteins from a mixture containing a non-naturally-occurring fusion protein according to claim 1 and impurities comprising the steps of:

selectively contacting said fusion protein with material which immobilized said fusion protein;

removing said impurities;

separating said fusion proteins from said material;

combining said fusion protein with DPP IV; and

isolating said desired protein.

14. A method according to claim 13 wherein said material is fixed in a column.

15. A method according to claim 13 wherein said material is an antibody which binds to said extension portion.

16. A method according to claim 13 wherein said material is immobilized metal ions and said extension portion comprises at least 3 consecutive X′ residues that are histidines.

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 91/09152

I. CLASSIFICATION OF SUBJECT MATTER (If several classification symbols apply, indicate all)[6]

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/62; C07K3/18; A61K37/43; C07K13/00
C12P21/02; // C12N15/16; A61K37/24

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C12N ; C07K ; A61K ; C12P

Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category °	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
X,P	WO,A,9 015 821 (THE UPJOHN COMPANY) 27 December 1990 cited in the application see the whole document, and especially : page 6, line 36 - page 7, line 16, and page 11, lines 4-16 ---	1,3-6, 10,12-13
X	EP,A,0 184 355 (ELI LILLY AND COMPANY) 11 June 1986	1,3-4
A	see page 3, line 16 - page 6, line 22 see page 18; table I & US,A,4 569 794 (ELI LILLY AND COMPANY) 11 February 1986 cited in the application --- -/--	2,7-9, 13-14,16

° Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
30 MARCH 1992	14.04.92
International Searching Authority	**Signature of Authorized Officer**
EUR PEAN PATENT OFFICE	ANDRES S.M.

2

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)		
Category °	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
A	EP,A,0 282 042 (F.HOFFMANN-LA ROCHE & CO., AG) 14 September 1988 cited in the application see column 1, line 53 - column 2, line 33; claims	1-16
A	EP,A,0 150 126 (IMMUNEX CORPORATION) 31 July 1985 see the whole document & US,A,4 782 137 (IMMUNEX CORPORATION) 1 November 1988 cited in the application	13-15
P,X	WO,A,9 115 589 (THE UPJOHN COMPANY) 17 October 1991 cited in the application see page 8, line 7 - page 13, line 6; claims	1,3-4, 6-9, 13-14,16

ANNEX TO THE INTERNATIONAL SEARCH REPORT ON INTERNATIONAL PATENT APPLICATION NO. US 9109152 SA 54941

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information. 30/03/92

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9015821	27-12-90	AU-A- 5674390	08-01-91
EP-A-0184355	11-06-86	US-A- 4569794	11-02-86
		AU-B- 588840	28-09-89
		AU-A- 5024085	12-06-86
		CA-A- 1252948	18-04-89
		JP-A- 61148197	05-07-86
US-A-4569794	11-02-86	AU-B- 588840	28-09-89
		AU-A- 5024085	12-06-86
		CA-A- 1252948	18-04-89
		EP-A,B 0184355	11-06-86
		JP-A- 61148197	05-07-86
EP-A-0282042	14-09-88	AU-B- 609783	09-05-91
		AU-A- 1270988	15-09-88
		JP-A- 63251095	18-10-88
		ZA-A- 8801534	12-09-88
EP-A-0150126	31-07-85	US-A- 4703004	27-10-87
		DE-A- 3584260	07-11-91
		JP-A- 60222429	07-11-85
		US-A- 4782137	01-11-88
US-A-4782137	01-11-88	US-A- 4703004	27-10-87
		DE-A- 3584260	07-11-91
		EP-A,B 0150126	31-07-85
		JP-A- 60222429	07-11-85
WO-A-9115589	17-10-91	AU-A- 7685791	30-10-91

EPO FORM P0479

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82